

Binding and segmentation of visual images by means of oscillatory neurons

M. Ursino, G.E. La Cara, and A. Sarti

Department of Electronics, Computer Science and Systems, University of Bologna, Bologna, Italy

Abstract—A neural network based on Wilson–Cowan oscillators is used to perform object recognition in a two-dimensional visual scene. The temporal correlation among groups of oscillating neurons is used as the main criterion to solve the classic binding and segmentation problem. The network uses an original pattern of short-range lateral excitations among adjacent neurons to achieve the binding problem, and an external inhibitory global neuron to provide segmentation of multiple objects in the same visual scene. The latter may represent an "attention mechanism" from neurons at a higher hierarchical level. Simulations performed by using multiple idealized figures (up to 4–5) in the presence of noise suggest that the network can satisfactorily recognize objects in most cases. However, the threshold and time constant of the attention mechanism depend on the complexity (number of objects and level of noise) of the scene under examination. The present results may be useful to improve our understanding of how distributed activities are integrated in the neural system to form single object perceptions. In perspective, the proposed model may find applications in practical algorithms for object recognition.

Keywords – Image segmentation, oscillatory neurons, object recognition.

I. INTRODUCTION

A fundamental task that the brain ordinarily solves in daily life is the segmentation of the visual scene into a set of distinct objects. This task requires the simultaneous solution of two complementary problems: a "binding" problem, which consists in assembling the common attributes of a single object into a unique figure, and the "segmentation" problem, which requires the visual scene to be decomposed in distinct figures, avoiding attributes of different objects to be grouped together. It is generally assumed that binding and segmentation of the visual scene is based on the principles of Gestalt psychology, such as proximity, similarity, common fate, connectedness, good continuation, etc.. [1], [2], [3], [4].

Despite the fact that the brain easily solves the binding and segmentation problem, its theoretical solution is still arduous using artificial neural networks. A traditional hypothesis is that information carrying different attributes of a same object converges to neurons at a higher hierarchical level. These neurons, in turn, respond selectively only to those groups of features which characterize a single object (grandmother cell representation). This assumption, however, involves many theoretical and neurophysiological problems, and it is usually rejected today.

A second, recent theory assumes that binding and segmentation are accomplished by the brain on the basis of temporal correlation between neural activities. Accordingly, neurons that fire in phase would signal common attributes of the same object. This hypothesis is supported by experiments in anesthetized cats and monkeys [1], [2]. In these studies stimuli which, according to the Gestalt theory, would belong to the same figure, cause synchronized activity in groups of neurons. Conversely, stimuli that belong to different figures

fail to induce synchronized responses.

In order to explore the previous aspects theoretically, a few models of oscillating neurons have been used in recent years [3], [4], [5], [6], [7]. These models suggest that binding and segmentation can be achieved using lateral connections between groups of oscillating neurons. However, many problems, especially concerning segmentation of multiple objects in the same scene, are still existent.

Aim of this work is to use an original neural network, based on Wilson–Cowan oscillators, to analyze the binding and segmentation problem in a two-dimensional visual scene in presence of noise. Original lateral connections are used to impose synchronism between neurons, while an attention mechanism is proposed to achieve segmentation. A few examples are presented and discussed.

II. SYSTEM DESCRIPTION

The model of a single oscillator consists of a feedback loop between an excitatory unit x_{ij} and an inhibitory unit y_{ij} . The time derivatives are defined as:

$$\begin{cases} \frac{d}{dt}x_{ij}(t) &= -x_{ij}(t) + H(x_{ij}(t) - \beta \cdot y_{ij}(t) + S_{ij} + I_{ij} + \varrho - \varphi_x - z(t)) \\ \frac{d}{dt}y_{ij}(t) &= -\gamma \cdot y_{ij}(t) + H(\alpha \cdot x_{ij}(t) - \varphi_y) + S_{ij} \end{cases} \quad (1)$$

where i and j represent the position of the oscillator within a two-dimensional network ($1 \leq i \leq N; 1 \leq j \leq M$), and:

$$H(\nu) = \frac{1}{1 + e^{-\frac{\nu}{T}}} \quad (2)$$

Equations (1), (2) describe essentially a simplified Wilson–Cowan oscillator. This oscillator model can be biologically interpreted as a mean field approximation of a network of excitatory and inhibitory neurons. The parameters have the following meaning: α and β are positive parameters describing the coupling between two units, particularly α influences the amplitude of oscillations; I represents external stimulation; ϱ denotes a noise term. $H(\nu)$ is a sigmoid activation function with thresholds φ_x and φ_y , for excitatory and inhibitory unit respectively. T affects the central slope of the sigmoidal relationship, and γ is inversely proportional to the time constant of the inhibitory units, hence it controls the frequency of oscillations. $z(t)$ represents the activity of a *Global Separator GS* which will be specified later on. S_{ij} represents

Report Documentation Page

Report Date 25 Oct 2001	Report Type N/A	Dates Covered (from... to) -
Title and Subtitle Binding and Segmentation of Visual Images by Means of Oscillatory Neurons		Contract Number
		Grant Number
		Program Element Number
Author(s)		Project Number
		Task Number
		Work Unit Number
Performing Organization Name(s) and Address(es) Department of Electronics Computer Science and Systems University of Bologna Bologna, Italy		Performing Organization Report Number
Sponsoring/Monitoring Agency Name(s) and Address(es) US Army Research, Development & Standardization Group (UK) PSC 802 Box 15 FPO AE 09499-1500		Sponsor/Monitor's Acronym(s)
		Sponsor/Monitor's Report Number(s)
Distribution/Availability Statement Approved for public release, distribution unlimited		
Supplementary Notes Papers from the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, October 25-28, 2001, held in Istanbul, Turkey. See also ADM001351 for entire conference on cd-rom., The original document contains color images.		
Abstract		
Subject Terms		
Report Classification unclassified		Classification of this page unclassified
Classification of Abstract unclassified		Limitation of Abstract UU
Number of Pages 4		

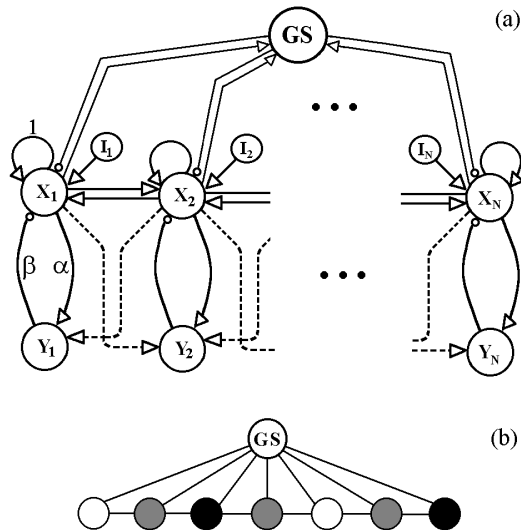


Fig. 1. (a) A diagram showing the interactions between the excitatory and the inhibitory units in a chain of oscillators. (Δ) represent excitatory connections; (\circ) represent inhibitory connections. Each excitatory unit is connected to the excitatory unit (solid line) and to the inhibitory unit (dashed line) of each nearest-neighbor oscillator. (b) Schematic representation: the neural oscillators in black circles are connected only to neural oscillators in grey circles. *GS* represents the *Global Separator*.

a term of coupling that we define as:

$$S_{ij} = 8 \cdot W \cdot \frac{\sum_{h=-1}^1 \sum_{k=-1}^1 F_{ijhk} x_{(i+h)(j+k)}}{\sum_{h=-1}^1 \sum_{k=-1}^1 F_{ijhk}} \quad (3)$$

where:

$$F_{ijhk} = \begin{cases} 0 & \text{if } h = k = 0 \\ & \text{or } i + h > N \quad \text{or } i + h < 1 \\ & \text{or } j + k > M \quad \text{or } j + k < 1 \\ 1 & \text{otherwise} \end{cases} \quad (4)$$

F_{ijhk} is a normalization factor for the connection weight W : in (3) the connections weights are opportunely normalized in relation to the oscillator position, which determines the number of neurons involved in coupling. In this way, the model respects an isotropy property (fig.1, fig.2). The terms S_{ij} in $H(\nu)$ implement the coupling between the excitatory unit of an oscillator and the excitatory units of its nearest-neighbors. The same terms S_{ij} in the equation of inhibitory unit implement the coupling between the excitatory unit and the inhibitory units of its nearest-neighbors: both coupling are excitatory. The second type of coupling, which was not used in previous models, is essential to improve synchronization of neurons within a same object.

The local excitatory connections assumed in our model conform with various lateral connections in the brain, in particular they could be interpreted as the horizontal connections in the visual cortex. Our simulations have revealed that, although the lateral excitatory couplings allow fast synchronization of all oscillators excited by a same stimulus, they do not permit satisfactory desynchronization of oscillators excited by different objects. For this reason we introduced a

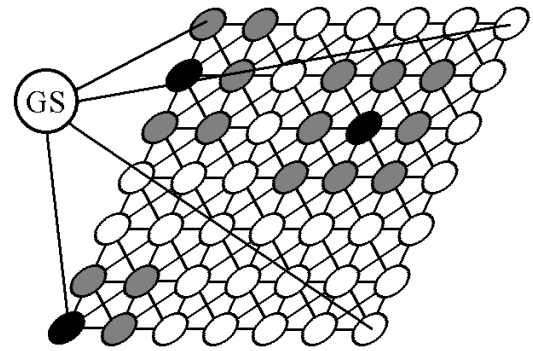


Fig. 2. Architecture of a two-dimensional network with eight nearest-neighbor excitatory coupling. An oscillator is indicated by a circle. The circle labeled *GS* represents the *Global Separator*. Three situations of connection are shown: each black circle is connected only to adjacent grey circles.

Global Separator GS through which we want to simulate a *mechanism of attention*. *GS* is an inhibitory interneuron defined as:

$$\frac{d}{dt} z(t) = \varphi \cdot (\sigma - z(t)) - \delta \cdot z(t). \quad (5)$$

where:

$$\sigma = \begin{cases} 1 & \text{if } \sum_{i=1}^N \sum_{j=1}^M x_{ij} \geq \vartheta \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

σ is a binary value and ϑ controls the level of activity of the entire network. The positive parameters δ and φ control the rate of growth and decay of $z(t)$, therefore the *segmentation capacity* of *GS*. We can speculate that σ , δ and φ fix the *degree of attention* of *GS*. *GS* receives excitatory input from the entire network and sends inhibition to all oscillators: this *long-connections* give rise to desynchronization (fig.1, fig.2). The inhibitory input is generated whenever in the neural grid there are some fairly active regions: only the neurons with enough activity will survive to inhibition continuing to oscillate.

Psychologically we know that the thalamic reticular complex may be involved in the global control of selective attention: it receives input from and sends projection to almost the entire cortex. The activity of *GS* should be interpreted as the collective behavior of the neural group in the thalamic reticular complex.

III. SIMULATION RESULTS

To illustrate how our network is used for image segmentation we have simulated a 15x15 and a 15x20 grid of neural oscillators with a Global Separator. In the first simulation we map two objects (designed as the sun and a car): for all stimulated oscillators $I=0.8$, for the others $I=0$. The image has been corrupted by means of an uniformly-distributed random noise. For the background the uniformly-distribution has mean equal to 0.1, while for the stimulated oscillators has mean equal to 0. The variance is equal to 0.003. The set of ordinary differential equations has been numerically solved on PENTIUM-based personal computers, using the fourth-order Runge-Kutta integration method with random initial conditions.

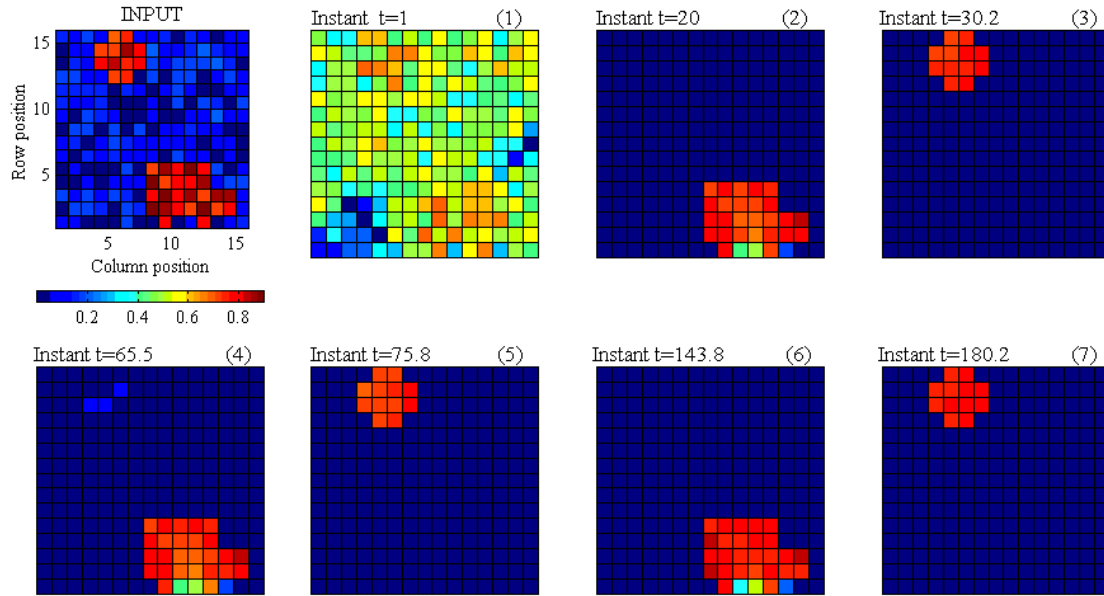


Fig. 3. The first picture represents the noisy input, on a noisy background, used for the network. The image is mapped to a 15X15 network and each red square denotes a single oscillator that receives input. Each following picture represents network activity at a time step in the numerical simulation. We name the objects as follows: a small car, a sun. The parameters are: $\alpha = 0.25$, $\beta = 2.5$, $\varphi_x = 0.7$, $\varphi_y = 0.15$, $\gamma = 1$, $W=1$, $I_{ij} = 0.8$, $T=0.025$ and for *GS*: $\vartheta = 1.8$, $\varphi = 2$, $\delta = 4.5$. The simulation took 2000 integration step.

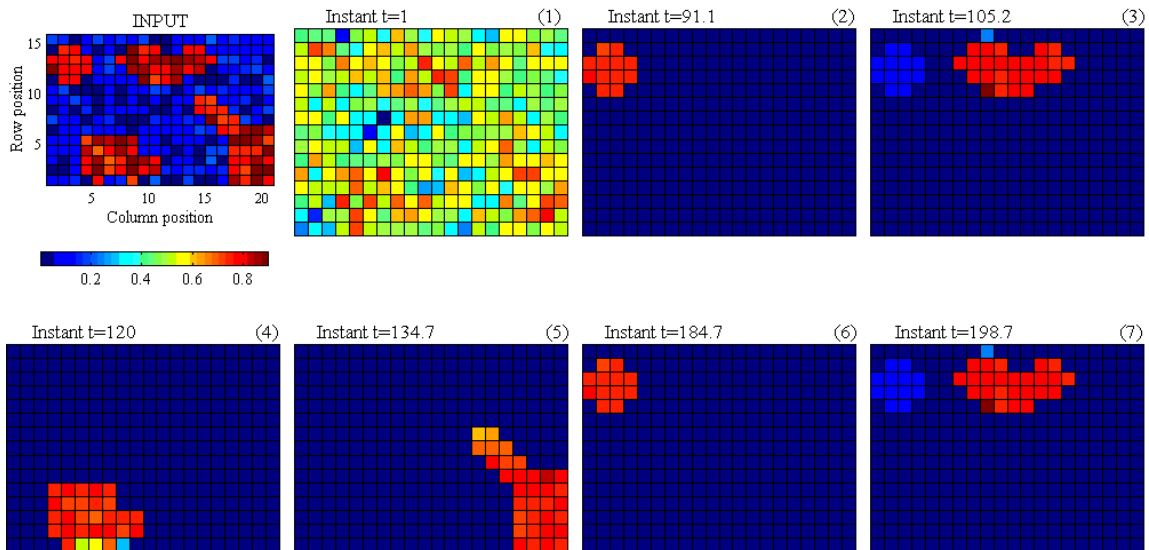


Fig. 4. A noisy image, on a noisy background, composed of four patterns mapped to a 15X20 network. The objects composing the image represent a small car, a garage, a sun and a cloud. The first picture represents the noisy input and each following picture represents network activity at a time step in the numerical simulation. The parameters for *GS* are: $\vartheta = 2$, $\varphi = 2$, $\delta = 8$. The other parameters are as specified in the captions of fig.3 except $\gamma = 0.5$. The simulation took 2000 integration step.

Fig.3 shows the instantaneous activity (snapshot) of each neuron of the network at various stages of dynamic evolution. A short time after the input is applied, we can observe the clear effect of synchronization and desynchronization: the noisy sun is segmented from the noisy car and from noisy background; then each object "fires" in a periodic way, separately from the others. Most parameters in (1), (2), (5), (6) are intrinsic to neural network and need not to be changed after they have been appropriately chosen. Only the parameters concerning *GS* and the oscillation frequency (γ) need to

be tuned for applications. In fact, with a fixed set of parameters, the dynamical system can segment only a limited number of patterns. This number depends on the ratio between the time that a single oscillator spends in the silent phases and the time that it spends in the active phases (segmentation capacity). Furthermore *GS* requires different degrees of attention. More particularly, in order to separate correctly the entire image, during each oscillation period of the network *GS* must be able to generate as many inhibitory impulses as the number of objects to be separated. To illustrate this point,

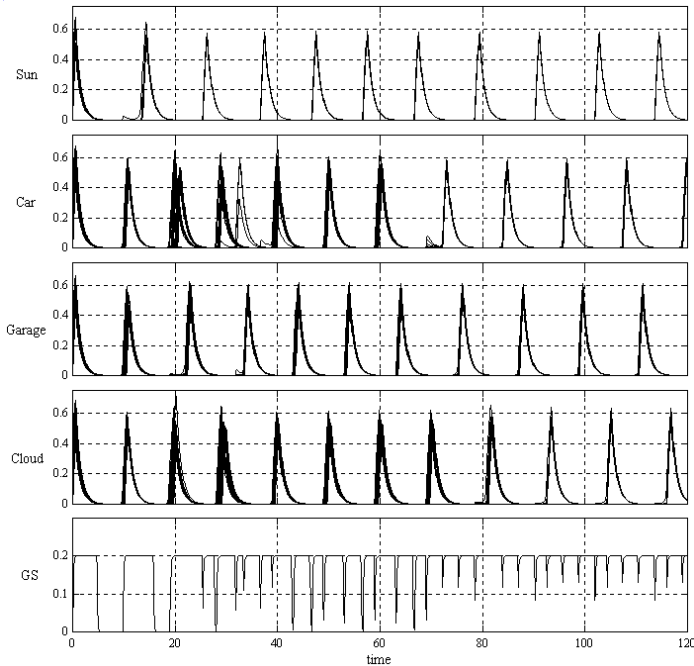


Fig. 5. Temporal evolution of every stimulated oscillator in the simulation of fig.4. The upper four traces show the combined x activities of four oscillators blocks representing the four corresponding objects indicated by their respective labels. The fifth trace shows the activity of the *Global Separator*.

we stimulated a 15×20 network with an arbitrary image containing four objects designed as a car, a garage, a cloud and the sun (fig.4). In this case for the correct segmentation we had to use a greater degree of attention for *GS* just because the image is more complex: the parameters in simulations of fig.3 and fig.4 are the same except for γ and GS^1 . The necessity to modify the parameters of the *GS* was also observed by Wang and Terman [3], although these authors used a different mechanism for global inhibition. Fig.5 shows the temporal evolution of the oscillators stimulated by each noisy object. The four upper traces represent the activities of four oscillators blocks and the bottom trace represents the activity of *GS*. The synchrony within each block and desynchrony between different blocks are achieved after a few cycles.

IV. DISCUSSION

The present study introduces some new aspects compared with previous models: i) we used an original pattern of lateral connections among groups of oscillating neurons. In particular, the presence of a short-range excitatory connection between the excitatory neurons, x_{ij} , and the adjacent inhibitory neurons, y_{ij} , allowed the attainment of robust synchronization in a large variety of visual scenes, even in the presence of noise. Without this connection, synchronization can be achieved only with difficulty in many cases. ii) Segmentation, in the presence of multiple objects, cannot be achieved with the use of short-range lateral connections only, but it requires the presence of an external inhibitory neuron. The same conclusion was reached by Wang et al too [3] through the use of a different mechanism of global inhibition. The external inhibitory neuron sends an inhibitory signal to all neurons

in the network, as soon as their global activity overcomes a given threshold. This global inhibition, in turn, allows only neurons belonging to a single object (i.e., the object with present maximal excitation) to fire, thus realizing segmentation in the same scene. It is interesting to observe that the global inhibition proposed in this work can simulate an "attention mechanism", originating from neurons in the cerebral cortex at a higher hierarchical level. Accordingly, we observed that segmentation of a different number of objects, and/or the use of a different noise level requires a modification in the threshold and time constant of the attention mechanism. This result agrees with the idea that a complex, noised visual scene asks for more attention to be correctly perceived, while lower attention is requested in the perception of a few objects in a noise-free ambient. At any given level of attention, just a limited number of distinct objects can be separately identified. This important property of the system agrees with the well-known psychological principle that there are fundamental limits on the number of simultaneously perceived objects. iii) An important feature of the present system is the capacity to recognize objects even in the presence of a strong noise superimposed on the visual scene. In particular, we observed that noise superimposed on the background causes only a mild deterioration in the system performance. When noise is superimposed directly on the object, the system can still solve the binding and segmentation problem in most cases, even when 4 or 5 objects are simultaneously present.

V. CONCLUSIONS

In conclusion, the present work suggests that the binding and segmentation problem can be performed by using temporal correlation among neuron activities. The solution of the problem benefits from short-range lateral connections among groups of oscillating neurons, and from the presence of an external attention mechanism. The mathematical form of both mechanisms is original compared with previous studies. The present results may be useful to improve our understanding of how distributed activities are integrated in the neural system to form single object perceptions. Moreover, the proposed model may find applications in practical algorithms for object recognition.

REFERENCES

- [1] W. Singer. Synchronization of cortical activity and its putative role in information processing and learning. *Annu. Rev. Physiol.*, 55:349–374, 1993.
- [2] W. Singer, C. Gray, A. Engel, P. Konig, A. Artola, and S. Brocher. Formation of cortical cell assemblies. In *Cold Spring Harbor Symposium on Quantitative Biology*, volume LV, 1990.
- [3] D. L. Wang and D. Terman. Image segmentation based on oscillatory correlation. *Neural Computation* (9) 805–836, 1997.
- [4] X. Liu and D. L. Wang. Perceptual organization based on temporal dynamics. *IEEE transactions on neural network*, pages 2679–2684, 1999.
- [5] S. Campbell and D. Wang. Synchronization and desynchronization in a network of locally coupled Wilson-Cowan oscillators. *IEEE transactions on neural network*, 7(3):541–554, 1996.
- [6] P. Konig and T.B. Schillen. Stimulus-dependent assembly formation of oscillatory responses: I. synchronization. *Neural Computation*, (3):155–166, 1991.
- [7] P. Konig and T.B. Schillen. Stimulus-dependent assembly formation of oscillatory responses: II. desynchronization. *Neural Computation*, (3):167–178, 1991.

¹For the parameter values see the captions of fig.3 and fig.4